DOCUMENT RESUME

ED 186 846 CS 005 413

AUTHOR Johnson, Dale D.; And Others
TITLE A New Approach to the Assessment of Phonics Skills.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Individualized Schooling.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO WRDCIS-TR-532
PUB DATE Dec 79
GRANT OB-NIE-G-80-0117
NOTE 99p.

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Beginning Reading; Elementary Education; Error Patterns; *Phonics; Primary Education; *Reading Diagnosis; *Reading Research; *Reading Tests; *Test Construction; Test Items

ABSTRACT

Work on the development of a phonics test that will be incorporated into a total word identification test battery has continued for several years. A prototype of the test was developed in winter 1977, and a revised version was administered in spring 1978. Analysis of this test data revealed several problems, and new criteria for the test were established. After a survey of the phonics components of nine popularly used diagnostic and achievement tests was conducted, a new test was designed to meet the new criteria. The new test assesses 45 different spelling-to-sound correspondences in 141 items and is comprised of two subtests on consonants and vowels. The subtests have been administered separately to small groups of students. Interesting patterns were noticed in the errors of students on the vowels subtest. Plans are being made to further revise the phonics test by increasing the number of letter-sound correspondences, and the test will be administered to younger students. (Author/MKM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED186846

Technical Report No. 532

A NEW APPROACH TO THE ASSESSMENT OF PHONICS SKILLS

by

Dale D. Johnson, Linda K. Shriberg,
Susan D. Pittelman, and Judy Schwenker

Report from the Project on
Studies in Language: Reading and Communication

Dale D. Johnson
Faculty Associate

Wisconsin Research and Development Center
for Individualized Schooling
The University of Wisconsin
Madison, Wisconsin

December 1979

CS005413

2

MAY 27 1980

Published by the Wisconsin Research and Development Center for Individualized Schooling. The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Center Grant No. OB-NIE-G-80-0117

MISSION STATEMENT

The mission of the Wisconsin Research and Development Center is to improve the quality of education by addressing the full range of issues and problems related to individualized schooling. Teaching, learning, and the problems of individualization are given concurrent attention in the Center's efforts to discover processes and develop strategies and materials for use in the schools. The Center pursues its mission by

- conducting and synthesizing research to clarify the processes of school-age children's learning and development
- conducting and synthesizing research to clarify effective approaches to teaching students basic skills and concepts
- developing and demonstrating improved instructional strategies, processes, and materials for students, teachers, and school administrators
- providing assistance to educators which helps transfer the outcomes of research and development to improved practice in local schools and teacher education institutions

The Wisconsin Research and Development Center is supported with funds from the National Institute of Education and the University of Wisconsin.

WISCONSIN RESEARCH AND DEVELOPMENT
CENTER FOR INDIVIDUALIZED SCHOOLING

Acknowledgments

The authors gratefully acknowledge the cooperation of the pupils and teachers at Oregon Elementary School and Oregon Middle School in Oregon, Wisconsin. Their support of and participation in this research is greatly appreciated. The authors also acknowledge with gratitude Lawrence D. Shriberg, Professor of Speech Pathology, University of Wisconsin-Madison; Dominic W. Massaro, Professor of Psychology, University of Wisconsin-Madison; Peter A. Schreiber, Professor of English and Linguistics, University of Wisconsin-Madison; and Richard L. Venezky, Unidel Professor of Educational Foundations, University of Delaware, for their helpful suggestions.

Table of Contents

Page

Acknowledgments iv

List of Tables. vii

List of Figures ix

Abstract. xi

Development of the Consonants Subtest 11

Method 11
Subjects. 11
Stimuli 11
Procedure. 21
Results. 23
Discussion 35

Development of the Vowels Subtest 38

Method 38
Subjects. 38
Stimuli 38
Procedure. 46
Results. 49
Discussion 64

Summary and Future Directions 66

Reference Notes 75

References. 77

Appendices

A A Page from the Picture Naming Booklet of the Consonants Subtest. 79

B The Subject's Copy of the First Page of the Word List from the Oral Reading Task for the Consonants Subtest; The Subject's Copy of the Poem from the Oral Reading Task for the Consonants Subtest. 81

C A Page from the Picture Naming Booklet for the Vowels Subtest. 85

D The Subject's Copy of the First Page of the Word List from the Oral Reading Task for the Vowels Subtest; The Subject's Copy of the Poem from the Oral Reading Task for the Vowels Subtest. 87

List of Tables

Table		Page
1	Consonants Subtest: Subject Population by School and Grade	12
2	Five Single-letter Consonant Correspondences Chosen from Three Frequency Groups	16
3	Consonants Subtest: Percent Correct Summary Statistics by Grade	25
4	Consonants Subtest: Comparison of Consonants in Initial and Final Position	26
5	Performance (% correct) by Grade on Single-letter Consonants, Consonant Clusters, and Consonant Digraphs	27
6	Item Analysis of Six Variant Consonant Correspondences for Total Subjects	29
7	Item Analysis of Digraph Correspondences	32
8	Summary Data on Oral Reading Tasks and Consonant Subtest Across Grades	34
9	Vowels Subtest: Subject Population by School and Grade	39
10	Vowels Subtest: Percent Correct Summary Statistics by Grade	50
11	Vowels Subtest: t-Tests for Performance Differences Due to Sex and Grade	52
12	Rank Ordered List of Performance Within Categories for the 20 Vowel Correspondences	53
13	t-Tests for Performances Differences Due to Vowel Category	55
14	Error Analysis by Vowel Category and Foil Type	59
15	Summary Data on Oral Reading Tasks and Vowels Subtest Across Grades	63
16	Comparison Between the Spring 1979 Phonics Test and the Proposed Revised Phonics Test for Number of Correspondences Assessed and Number of Items per Category	73

List of Figures

Figure		Page
1	The directions and practice items from the first page of the Consonants Subtest.	14
2	The directions and practice items from the first page of the Vowels Subtest.	40
3	Item analyses of the three items assessing short o showing percent of subjects selecting responses	56

Abstract

The research discussed in this report continues an effort to identify subskills of word identification which correlate highly with reading comprehension and to develop empirically based instruments to assess these subskills.

A prototype of the Phonics Test was developed in winter 1977 and a revised version was administered in spring 1978. Analyses of the test data revealed several problems inherent in the instrument. As a result, the criteria which had guided test construction were evaluated, and new criteria were established for developing the present test. In addition, a survey of the phonics components of nine popularly used diagnostic and achievement tests was conducted in an effort to identify any existing phonics instrument that would meet the established criteria. The survey revealed there is no instrument currently available that takes into consideration all of the criteria essential to phonics assessment. Therefore, a new test was designed to meet these criteria.

The new Phonics Test assesses 45 different spelling-to-sound correspondences in 141 items. The test is comprised of two subtests: A Consonants Subtest which assesses 25 correspondences (single-letter consonants, consonant clusters, and consonant digraphs); and a Vowels Subtest which assesses 20 correspondences (long vowels, short vowels, and vowel clusters).

The Consonants Subtest was administered in winter 1979 to 187 third, fourth, and fifth grade students. A randomly selected subsample

of 59 students also participated in an oral reading task, to establish concurrent validity for the new test. Results of this study indicated that by third grade, children have mastered most of the frequently occurring letter-sound correspondences in our language. When children did make errors, they usually made them on either consonant digraphs or on the three single-letter consonants which have other common sound correspondences (c, g, and s).

The Vowels Subtest was administered in spring 1979 to 190 third, fourth, and fifth grade students. A randomly selected subsample of 42 students also participated in an oral reading task. The results of the study indicated that, overall, performance was highest on vowel clusters, next highest on long vowels, and lowest on short vowels. The most interesting result of the study, however, was when children did make errors, they appeared to have been strategic in their responses. Most children who chose an incorrect response choice tended to select the response choice that had the long vowel counterpart to the short vowel in the target correspondence (or vice versa). In the case of vowel clusters, students tended to select the response choice that had the sound of one of the two letters of the target cluster.

The item analyses from the winter and spring 1979 Phonics Test raised some interesting speculations about the nature of phonics processing. Plans are being made to further revise the Phonics Test and to administer the new instrument in winter 1979-80. To make the revised instrument more global, a greater number of letter-sound correspondences will be assessed. Also, because performance on the Consonants Subtest and the Vowels Subtest was extremely high, younger

students (second and third graders) will participate in the winter 1979-80 test administration.

The final version of the Phonics Test will be incorporated into a total Word Identification Test battery. We expect that the information gained through the administration of this battery will provide educators with insights about the relationships of the various components of word identification to comprehension. This knowledge will have its ultimate benefit in the classroom, by providing teachers with a valid and reliable instrument for assessing reading skills and for planning reading instruction.

The focus of the research conducted over the last 3 years by the Project on Identification and Assessment of Components of Reading Comprehension has been to (1) explore the relationships among word identification skills and comprehension abilities, (2) determine to what degree and in what ways the mastery of skills relates to reading comprehension, and (3) identify those skills which appear to relate most closely to the comprehension of written language.

Because much of the current information concerning word identification and comprehension is based on speculation only, an empirical examination of these issues is clearly needed. A primary goal of the Project, therefore, has been to investigate the possible relationships among word identification skills and reading comprehension, and to ascertain how these skills might be assessed most efficiently.

Beginning in winter 1977, a test battery was developed to assess skills in the three broad areas of word identification: phonics, structure, and context. During spring 1978, the Project-developed Word Identification Test battery, and the reading subtest of the Metropolitan Achievement Tests (Durost, Bixler, Wrightstone, Prescott, & Balow, 1970), were administered to over 1,400 public elementary school children in grades two, four, and six from five regions of the country (Johnson, Pittelman, Schwenker, Shriberg, & Morgan-Janty, 1978). The Phonics component consisted of two subtests: a 20-item Consonants Subtest and a 20-item Vowels Subtest. For each item, a target synthetic word appeared in a box and had one, two, or three letters underlined. The children's task was to circle the correct

response (out of three real-word choices) "whose underlined part sound[ed] just like the underlined part of the word in the box."

Subsequent data analyses revealed poor item-subtest reliability for items where (1) single-letter consonants and consonant digraphs were tested in the medial position in words; (2) medial vowels were tested in words ending with liquids (r or l); (3) similar vowel sounds in words could be perceived as the same vowel sound by users of different dialects; and (4) the underlined part of a target synthetic word had a different word position (i.e., initial, medial, or final) from the underlined part of one or more response choices. The spring 1978 Phonics Test also fell short on validity measures, especially because it did not control for the visual matching of letters as a basis for selecting a response choice. Plans were made, therefore, to redevelop the Phonics Test and to administer the revised test to new populations of elementary school children.

Several issues have to be considered in the development of a phonics instrument. First, the scope of the test has to be addressed, that is, a decision has to be made as to which of the hundreds of spelling-to-sound correspondences in our language should be selected for assessment. Next, the format of the test has to be considered. It is important that the test mirror the reading process as closely as possible, requiring children to use a decoding (from letter to sound), rather than an encoding (from sound to spelling) process. While an individually administered productive task would best reflect the reading process, consideration must be given to efficiency of assessment. Finally, the modes in which the target spelling-to-sound

correspondences and the response choices are presented must be determined.

In a phonics assessment instrument, there are several ways to present the target letter-sound correspondences in a decoding (rather than an encoding) format. The target stimulus could be the letter(s) in isolation, or the target stimulus could be presented within a real word or a synthetic word. Similarly, there are several considerations involved in developing the response choices. One concern would be the number of response choices; too few would reduce the reliability of the test and too many would make the test unwieldy. A second consideration is the mode in which the response choices appear: isolated letter(s), words, or pictures. If the response choice is a word or picture, attention must be given to the position of the targeted sound within the word or picture name (beginning, medial, or final), and to the number of syllables (one or more) that make up that word.

A survey of the phonics components of popularly used tests was conducted in light of the above issues. The following diagnostic and achievement tests were examined: the California Achievement Tests (McGraw-Hill, 1977), the Botel Reading Inventory (Botel, 1961), the Prescriptive Reading Inventory (McGraw-Hill, 1972, 1976), the Skills Monitoring System for Reading and Word Identification (Harcourt Brace Jovanovich, Inc., 1975), the Wisconsin Design for Reading Skill Development (Otto, Miles, Kamm, & Stewart, 1972), the Phonics Knowledge Survey (Durkin & Meshover, 1964), the California Phonics Survey (Brown & Cottrell, 1963), the Stanford Achievement Test (Madden, Gardner, Rudman, Karlsen, & Merwin, 1970-1974), and the Silent Reading

Diagnostic Tests (Bond, Balow, & Hoyt, 1970).

Many of the tests surveyed assess only a small number of the many correspondences in our language. The California Achievement Tests, for example, consist of only 25 items; 10 items assess the entire Consonants category (single-letter consonants, consonant digraphs, and consonant clusters), 13 items assess single-letter vowels (all either long or short), and 2 items assess vowel clusters or diphthongs. The Prescriptive Reading Inventory and the Phonics Knowledge Survey, both diagnostic tests, also have relatively few items which form the basis for global judgments regarding children's overall competence with phonics.

In addition to the number of correspondences assessed, attention must be focused on how often those correspondences occur in our language. According to the frequency data for the spelling-to-sound correspondences in the 20,000 most common English words (Venezky, Note 1) many of the tests we reviewed assessed correspondences that have low frequencies of occurrence. The California Achievement Tests, the Prescriptive Reading Inventory, and the Wisconsin Design all include a sizeable number of items that assess correspondences which appear relatively infrequently in our language. Moreover, a large number of correspondences are not assessed in the position(s) in which they most typically occur in English words. The Wisconsin Design, for example, assesses the single letter v in final word position. Data from the Venezky tabulations indicate that v appears only twice in final position in the 20,000 most common English words. One has to consider, therefore, whether it is educationally prudent to assess v

as /v/ in final position.

Clearly, the most accurate procedure by which to assess phonics skills is by means of a productive task. The ideal phonics test would require that the child read aloud, while the examiner records all pronunciation errors made on unfamiliar words. The Phonics Knowledge Survey is an example of a productive test, although it uses letters instead of words. Children view separately each of 14 consonants and the 5 single-letter vowels; they are then asked to pronounce the corresponding sounds for the consonant letters, and the 5 long and 5 short corresponding sounds for the vowel letters. This productive method of assessment is not efficient, however, because of the time needed for testing each child individually. In addition to the problems involved with the procedure, the validity of assessing letter-sound correspondences in isolation, rather than within the context of words, is also questionable.

The Botel Reading Inventory, also a productive test, utilizes a written format which can be administered to a large group or class. Unfortunately, however, some of the tasks on the Botel may not be appropriate for all the children for whom it is intended. Level B, for example, requires children to write a two-letter diphthong that has the same sound as a given single-letter vowel--a rather sophisticated task for primary-level pupils. Moreover, the Botel involves encoding as well as decoding processes.

Another issue central to the construction of an effective phonics instrument is the mode of presentation of the letter-sound correspondences being assessed. The letters for a correspondence can be

presented in isolation, in real target words, or in synthetic target words. The value of assessing the sounds of letters in isolation, however, is questionable, because the sounds of many letters, especially vowels, are determined by their orthographic environments.

One alternative to assessing letters in isolation is to present the target letters within a real word. But the problem inherent in using real words is that children may recognize the words as sight words, and therefore might not need to utilize a decoding strategy. The Skills Monitoring System for Reading and Word Identification, for example, assesses the letter-sound correspondence for ea as it appears in the real word each. The four response choices are dish, Christmas, anchor, and chair. Because all four response choices include accurate pronunciations for ch, children must rely on prior knowledge of the pronunciations of the target word and response choices to arrive at the correct answer. This implies that children must recognize these words as sight words, and the process is then one of auditory matching rather than of decoding.

A second alternative is to present the target letters for a correspondence within a synthetic word. This format allows letter-sound correspondences to be presented within appropriate orthographic environments, and requires that children use phonics (decoding) skills, rather than a sight word approach, to arrive at correct answers. One concern with the use of synthetic target words, however, is ensuring that the synthetic words conform to phonological principles of the English language. In the tests surveyed, the synthetic words were not always phonologically accurate. For example, in the Phonics Knowledge

Survey, children are asked to pronounce the sound made by the a in the synthetic target word, aef. The correct answer is given as long a, because children are expected to apply the "rule" that when two vowels are together, the first vowel says its name and the second vowel is silent. According to Venezky (1970), however, ae in initial position, is never pronounced as long a.

When creating a phonics test, it is also important that careful attention be given to the development of the response choices. One area of concern is the number of response choices because this can affect the reliability of the test. Many of the tests surveyed had true-false, same-different, or yes-no formats, which greatly increase the likelihood of guessing correct answers. Most of the tests, however, had a multiple choice format, with the number of response choices varying from three to five.

A second consideration affecting the development of response choices is the number of syllables in words. Many of the tests were not consistent in controlling for the number of syllables in the response choices within an item. The Wisconsin Design, the Skills Monitoring System for Reading and Word Identification, and the Stanford Achievement Test, for example, all include response choices with varying numbers of syllables. In this regard, educators have expressed concern that decoding a multisyllable word may require different underlying skills than decoding a one-syllable word, and that including both kinds of words within a given test item may confuse young children. The California Phonics Survey contains many items with multisyllable and one-syllable words, but because this test was developed for

secondary and college level pupils, this issue is of lesser concern here.

A third issue has to do with the position of the target letters of the correspondence within a word (initial, medial, or final). According to Venezky (Note 1), the position(s) of greatest occurrence varies for different letters: it seems logical, therefore, to assess a letter-sound correspondence in the word position where it most frequently occurs. The position of the target letters within a target word should also match the position of the foil correspondences within each of the response choices. The Stanford Achievement Test, the Skills Monitoring System for Reading and Word Identification, and the Prescriptive Reading Inventory, include items that contain shifts in the word position of target correspondences. The Prescriptive Reading Inventory, for example, has an item in which short a is presented in medial position of the CVC word, cat. One of the response choices is the CV word, day, which has its vowel sound in final position. Asking young children to perform these word position shifts may require a different psychological process than the one used when target correspondences are all in the same position within words.

A fourth issue related to the development of response choices is the mode in which these response choices appear. All but two of the tests reviewed (the Phonics Knowledge Survey and the Botel Reading Inventory are productive tests) include response choices in the form of isolated single letters or letter clusters, real words, or pictures.

The validity of utilizing single-letter or letter cluster response choices (as in the Prescriptive Reading Inventory and the Silent

Reading Diagnostic Tests) is questionable. In a review of the Silent Reading Diagnostic Tests, for example, Kress (1972) refers to the response choices as "artificial graphic representations," and is dubious about whether they truly measure the phonics abilities being assessed: ". . . in Test 6, for the beginning sound in natural, the child is to select pn, from pn, un, tn, nt; in Test 7, for the ending sound in decay, he is to select quet from khak, kayn, cove, quet; in Test 8, for the vowel sound at the beginning of the word win, he is to select i from x, i, v, a."

In most of the tests, the most common form of response choice is the real word (the Prescriptive Reading Inventory, the Skills Monitoring System for Reading and Word Identification, the California Phonics Survey, the California Achievement Tests, and the Stanford Achievement Test). One of the problems with using real word response choices is the likelihood of testing visual matching, rather than decoding. In the Stanford Achievement Test, for example, children are presented with the target letter t in the real word ten. The three response choices are gate, nine, and been. Children can easily select the correct answer by visually matching the t in ten with the t in gate.

Of all the tests inspected, the Wisconsin Design was most consistent in using response choices in picture form. The use of pictures as response choices eliminates the problems associated with real or synthetic words--that is, the visual matching of real words, the recognition of real words as sight words, and the concern that synthetic words may not conform to phonological properties of the English language. In the Wisconsin Design, the examiner pronounces the picture names of

the target word and of all the response choices. The focus of this test, therefore, is on the auditory matching of sounds, rather than on the decoding of letter-sound correspondences.

It is apparent from the review of existing tests, that a phonics instrument which considers all the issues discussed above is not currently available. As a result, the Project on Identification and Assessment of Components of Reading Comprehension decided to revise the spring 1978 Phonics Test. This test was redeveloped to use a decoding format and to include items for a large number of frequently occurring spelling-to-sound correspondences. The new test presents target correspondences in synthetic words that are phonologically accurate and four response choices in picture form. A format consisting of synthetic target words and pictorial response choices will ensure that the test assesses true phonics skills, rather than the visual matching of letters and the auditory matching of sounds.

The revised Phonics Test assesses 45 different spelling-to-sound correspondences with a total of 141 items. The Phonics Test is comprised of two subtests: a Consonants Subtest, which is composed of 51 single-letter consonant items, 15 consonant cluster items, and 15 consonant digraph items; and a Vowels Subtest, which is composed of 15 long vowel items, 15 short vowel items, and 30 vowel cluster items. Each spelling-to-sound correspondence is tested with three or four items. Selection of target items was based on frequency data from the Venezky (Note 1) collection of spelling-to-sound correspondences of the 20,000 most common English words. Response choices are based on speech production data and perceptual information from the Bouma (1971),

Miller and Nicely (1955), and Peterson and Barney (1952) confusion matrix studies.

With a continuing emphasis on the importance of phonics instruction in elementary school reading programs, educators are clearly in need of an effective instrument for assessing phonics. The goal of the Project, then, is to develop a phonics test which will ultimately be incorporated into a total word identification test battery, and which will be a valid and reliable instrument for both researchers and practitioners in the field of education.

DEVELOPMENT OF THE CONSONANTS SUBTEST

The Consonants Subtest assesses 25 different spelling-to-sound correspondences. Sounds selected for testing were single-letter consonants, consonant clusters, and consonant digraphs.

Method

Subjects

A total of 187 third, fourth, and fifth grade pupils from the Oregon, Wisconsin Public School District participated in the studies (see Table 1). Third grade pupils were from Oregon Elementary School; fourth and fifth grade pupils formed a mixed-class unit in Oregon Middle School. Oregon is a midwestern rural community with a socio-economic level ranging from lower middle to upper middle class.

Stimuli

The Consonants Subtest assessed 25 different spelling-to-sound

Table 1

Consonants Subtest:

Subject Population by School and Grade

(N = 187)

Schools and classrooms	Grade 3		Grade 4		Grade 5	
	Consonants Subtest	Oral reading tasks	Consonants Subtest	Oral reading tasks	Consonants Subtest	Oral reading tasks
Oregon Elementary School						
Classroom A	21	8				
Classroom B	26	9				
Classroom C	23	8				
Classroom D	19	7				
Oregon Middle School						
Classroom A			9	3	7	3
Classroom B			14	5	6	1
Classroom C			11	4	9	1
Classroom D			15	4	8	1
Classroom E			12	4	7	1
Totals	89	(32)	61	(20)	37	(7)

correspondences in 81 items. Each item consisted of a synthetic word with the target letter(s) underlined and four response choices in picture form. Pupils were instructed to read the synthetic word to themselves and decide the sound of the underlined letter(s). Next they were told to circle the picture in that row which had a name beginning (or ending) with the sound of the underlined letter(s). Figure 1 is a copy of the directions and practice items from the Consonants Subtest.

Target consonant sounds were selected according to their frequencies of occurrence in the Venezky (Note 1) corpus of the 20,000 most common English words. Only those spelling-to-sound correspondences with total frequencies of 150 or more are included in the Consonants Subtest. These correspondences appear in one-syllable, target synthetic words that conform to phonological properties of the English language.

The response choices for each item are four pictures whose names are well known to elementary school-age children. For each target item, the four response choice categories are: (1) a Correct response choice, (2) an Acoustically Close response choice, (3) a Visually Close response choice, and (4) a Neither Acoustically Close Nor Visually Close ("Neither") response choice. (For 12 of the single-letter consonant items, this fourth response choice category was changed to include the second most common sound correspondence for those consonants.)

Selection of Target Sounds

Single-letter consonants. According to the Venezky (Note 1) tabulations all 22 single-letter consonant correspondences have

PHONICS: Consonants
Initial Position

In each row, look at the made-up word. Notice that there is a letter or letters underlined at the beginning of that word. Read the word to yourself and decide how the underlined letter or letters sound. Then find the picture in the row whose name begins with the same sound as the underlined letter or letters. Draw a circle around that picture.

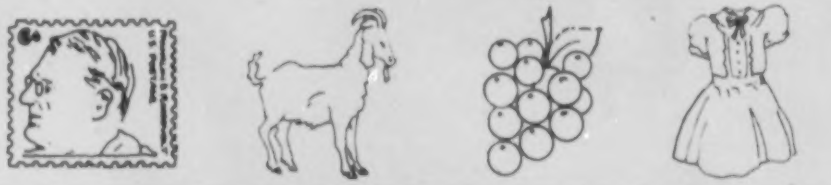
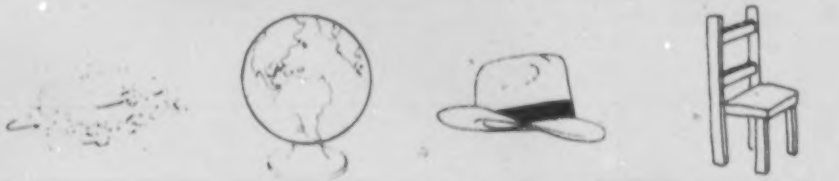
A. <u>f</u>ape	
B. <u>gr</u>oon	
C. <u>ch</u>ilp	

Figure 1. The directions and practice items from the first page of the Consonants Subtest.

frequencies of 150 or more in the 20,000 most common English words. These single-letter consonant correspondences were divided into three frequency groups and five were randomly chosen from each group. The 15 single-letter consonant correspondences are listed in Table 2.

Consonant clusters. Sixteen two-letter consonant clusters were identified as having frequencies of over 150. The following five consonant clusters were randomly selected for inclusion in the Phonics Test: st, tr, sp, br, and cl.

Consonant digraphs. Of the seven two-letter consonant digraphs, Venezky (Note 1) listed five with frequencies greater than 150. All five were selected for inclusion in the Phonics Test: sh, ng, th (voiceless), ch, and ph.

Single-letter consonants, consonant clusters, and consonant digraphs were tested in the initial or final positions of words, according to their frequencies of occurrence in those positions as reported in the Venezky tabulations. Some spelling-to-sound correspondences were tested in initial position only, some in final position only, and some in both initial and final positions. Consonants, consonant clusters, and consonant digraphs were not tested in medial position because (1) a medial consonant sound would have to be assessed in words that exceed one syllable, (2) a medial consonant sound might be "masked" by initial and final consonant sounds, and (3) consonant clusters often lose their integrity as a cluster in the medial position in words (e.g., when the word, master, is syllabicated into mas and ter, the cluster st becomes final s and initial t).

Table 2

Five Single-letter Consonant Correspondences Chosen from Three Frequency Groups

Frequency group	Target sounds selected
2000 - 5000 occurrences	n, t, d, c as /k/, s as /s/
700 - 2000 occurrences	p, b, v, g as /g/, c as /s/
below 700 occurrences	s as /z/, g as /j/, w, k, j

Selection of Response Choices

Each of the 81 items in the Consonants Subtest has 4 response choices in picture form: a Correct response choice, an Acoustically Close response choice, a Visually Close response choice, and a Neither Acoustically Close Nor Visually Close (Neither) response choice.

Correct response choice. This response choice is a picture whose name contains an initial or final consonant sound that is the same as the sound corresponding to the underlined letter(s) in the target synthetic word.

Acoustically Close response choice. This response choice is a picture whose name contains an initial or final consonant sound that is acoustically close to the sound corresponding to the underlined letter(s) in the target synthetic word. Whenever possible, data from the Miller and Nicely (1955) confusion matrices were used in selecting the acoustically close sound. (The Miller and Nicely study presented perceptual confusions for 16 consonant singletons under the manipulated variables of signal to noise ratio, decibel level, and frequency response in cycles per second.) For example, one of the test items for the spelling-to-sound correspondence for b as /b/ in initial position presents the target synthetic word bund. The Miller and Nicely study indicated that the sound /b/ is most often confused with the sounds of /v/ or /f/. The Acoustically Close response choice for this item, then, is vest, a picture whose name begins with the sound /v/. (Another Acoustically Close response choice might have been foot, because foot begins with /f/.)

The Miller and Nicely study did not provide confusion information

for some single-letter consonants, any consonant clusters, or some consonant digraphs. Therefore, speech production data (Massaro, 1975; Shriberg, Note 2) were also used to construct Acoustically Close response choices. (There is evidence to suggest that sounds produced in proximal positions within the oral cavity are more likely to be confused than sounds produced in more distal positions. An Acoustically Close response choice could therefore be determined from this information.)

For example, for the target consonant cluster st we decided to retain the fricative /s/, and to change the /t/ to another voiceless stop (hence: /sp/ or /sk/). The target synthetic word zoost appears in a test item for st in final position. The Correct response choice is a picture of a nest, because nest ends with the sound /st/, the same sound as the /st/ in zoost. The Acoustically Close response choice is a picture of a desk, because desk has the sound of /sk/ in final position. (Another Acoustically Close response choice might have been a picture of a wasp, because wasp ends in /sp/, the other Acoustically Close response choice possibility for /st/.)

Visually Close response choice. This response choice is a picture whose name contains an initial or final letter that is visually close to the underlined letter(s) in the target synthetic word. Data from Bouma's (1971) confusion matrix study of lower case letters were used to create the Visually Close response choices. (The Bouma study reported confusions made for each of the 26 lower case letters when presented on a screen in IBM Courier 10 typeface, 7 degrees from

the point of fixation at a distance of 50 centimeters.) To illustrate how the matrix information was used, consider the target letter t: Bouma's confusion matrix indicates that lower case t is most frequently confused with lower case i. In an item that tests the sound of t in initial position, therefore, a Visually Close response choice could be a picture of a cube of ice, because ice begins with i.

The Bouma confusion matrix was not applicable for two-letter clusters and digraphs. Instead, a picture was used whose name contained one of the two consonants of the cluster or digraph. In the item where st is tested in final position within the target synthetic word zoost, the Visually Close response choice is a picture of a bus, because bus ends in one of the letter(s) of the two-letter cluster st. (Another Visually Close response choice could have been a picture of a goat, because goat ends in t, the other of the two letters of the cluster.)

Neither Acoustically Close Nor Visually Close (Neither) response choice. This response choice is a picture whose name has neither a sound that is acoustically close nor letters that are visually close to the underlined letter(s) in the target synthetic word. For example, in the item testing final st in the target synthetic word zoost, the Neither response category is a picture of a worm. The final consonant cluster rm is neither acoustically close nor visually close to the st in zoost. The response categories were created to see whether children unable to decode a target correspondence would apply an auditory strategy (select the Acoustically Close response choice), a

visual strategy (select the Visually Close response choice), or if they would merely guess. A disproportionate number of incorrect responses in the Neither response choice category might indicate a tendency toward guessing.

Three single-letter consonants (c, s, and g) each have two common sound correspondences. In the items assessing these correspondences, the Neither response category was not always used. The six spelling-to-sound correspondences for these three consonants, and the positions in which they were assessed, are:

c as /k/	(initial)
c as /s/	(initial; final with silent e)
s as /s/	(initial; final)
s as /z/	(final)
g as /g/	(initial; final)
g as /j/	(final with silent e)

For each of the six correspondences above, four items (instead of three) were created. Of the four items, two had the response categories of Correct, Acoustically Close, Visually Close, and Neither. For the remaining two items, a response choice called Other Common Sound Correspondence was substituted for the Neither response choice. This new category contained a picture whose name began or ended with the other common sound correspondence. For example, in an item testing c as /k/ in initial position, the target synthetic word is cobe, the Correct response choice is a picture of a cuff, and the response choice in the Other Common Sound Correspondence category is a picture of a saw (the other common sound correspondence for c is /s/).

Procedure

It was important to ensure that the pictures in the Consonants Subtest be identified by their designated names. Therefore, each class was given a Picture Naming Exercise prior to the Consonants Subtest of the Phonics Test. Upon completion of the subtest, a random sample of students also participated in two oral reading tasks.

Picture Naming Exercise. The selection of pictures for the Picture Naming Exercise was based on a pilot study involving approximately 30 fourth grade pupils. All pictures that were incorrectly labeled by two or more subjects in the pilot study were included in the Picture Naming Exercise.

A two-page Picture Naming booklet consisting of 31 pictures (based on the results of the pilot study) was distributed to each student. The pictures were arranged in rows of four. Many of the rows were made up of pictures which shared a common feature--that is, a plural name or an arrow which pointed to a particular part of the picture. (Appendix A contains a page from the Picture Naming booklet for the Consonants Subtest.) Students were instructed to look at one row of pictures at a time. The examiner pronounced the name of each picture in that row and the children repeated the name aloud. When the names of all 31 pictures had been identified through this procedure, the Picture Naming booklets were collected.

Consonants Subtest. Following the Picture Naming Exercise, each student was given a copy of the Consonants Subtest. The examiner

explained to the children that they would be listening for consonant sounds at the beginning of words. The examiner then instructed the students in each of the three practice items. At the end of the practice period (approximately 9 minutes), the students were instructed to work independently on the actual test items until they reached the large stop sign on the eighth page. The children were told that if they completed all 49 items before the allocated time was up, they could go back and review their answers.

After completing the Initial Consonants section of the Consonants Subtest, the examiner explained to the children that they were going to be listening for consonant sounds at the end of words. The examiner instructed the students in the three practice items for final consonant sounds. At the end of the practice period (approximately 3 minutes), the children were directed to work independently on the actual test items and to complete all 32 items remaining in the booklet. The children were told that if they completed the Final Consonants section before the allocated time was up, they could review any of their work in the entire booklet.

The students took approximately 30 to 35 minutes to complete both the Initial Consonants and Final Consonants sections (including practice items) of the Consonants Subtest.

Oral reading tasks. Two oral reading tasks were individually administered to assess for concurrent validity of the Consonants Subtest. Approximately 32% of the children from each classroom were randomly selected to participate. The first task required that the student read aloud a list of 81 target synthetic words from the

Consonants Subtest. As the student read the list of words, the examiner noted all pronunciation errors made on target sounds.

Following the oral reading of the word list, the student was instructed to read aloud a 211-word poem. The poem included several words containing each of the spelling-to-sound correspondences assessed in the Consonants Subtest. The examiner noted all pronunciation errors on target correspondences. (The student's copy of the first page from the word list and the poem are presented in Appendix B.)

The total time for the administration of the oral reading tasks ranged from 2 to 15 minutes, with the word list taking 1 to 10 minutes and the poem taking 1 to 5 minutes. In general, third grade students required more time to perform the tasks than did fourth and fifth grade students.

All tests were administered by staff members from the Wisconsin Research and Development Center for Individualized Schooling. Directions for the Picture Naming Exercise and the Consonants Subtest were read from an Administrator's Manual. Testing was conducted on February 6, 1979.

Results

A total of 187 pupils in grades three, four, and five participated in the study to evaluate the Consonants Subtest. Summary statistics on the Consonants Subtest, by grade level, are presented in Table 3. Mean percent correct on the total subtest (81 items) was above 90% for all three grades. Overall Hoyt reliability estimates were .78,

Table 3

Consonants Subtest:

Percent Correct

Summary Statistics by Grade

Grade 3	Grade 4	Grade 5
$\underline{N}$ = 89	$\underline{N}$ = 61	$\underline{N}$ = 37
$\bar{X}$ = 91.27%	$\bar{X}$ = 92.27%	$\bar{X}$ = 93.29%
$\underline{SD}$ = 5.53	$\underline{SD}$ = 6.39	$\underline{SD}$ = 5.15
$Range_{\%}$ = 59.26 - 97.53	$Range_{\%}$ = 51.85 - 98.77	$Range_{\%}$ = 71.60 - 97.53
$Range_{RS}$ = 48 - 79[a]	$Range_{RS}$ = 42 - 80[a]	$Range_{RS}$ = 58 - 79[a]

[a]Out of a possible 81 items.

.86, and .81 for grades 3, 4, and 5, respectively. Item analyses were performed to gain information about individual items, as well as the item categories within the total test.

Three types of analyses were performed on the data: a comparison of mean scores for consonants in initial versus final position; a comparison of mean scores for single-letter consonants, consonant clusters, and consonant digraphs; and an examination of the mean scores for the three or four items assessing each of the 25 spelling-to-sound correspondences assessed in the Consonants Subtest.

The first analysis compared performance on items assessing consonants in initial position with performance on items assessing consonants in final position. As illustrated in Table 4, children at each grade level achieved higher performance on consonants in initial position than on consonants in final position. Mean differences between performance on initial versus final items averaged 4.48% across grades.

The second analysis compared performance on the three categories of consonants which comprised the total subtest: single-letter consonants (51 items), consonant clusters (15 items), and consonant digraphs (15 items). Results from this analysis are presented in Table 5. Once again, differences between grades were small. For all three grades, performance was highest on clusters and lowest on digraphs.

The third analysis examined performance on the 25 different spelling-to-sound correspondences assessed in the 181-item Consonants Subtest. Performance on all test items for 17 of the 25 correspondences

Table 4

Consonants Subtest: Comparison of Consonants in Initial and Final Positions

	Number of subjects	Mean % correct	SD	Total number of items	Range of scores (% correct)
Total group	187				
Initial		93.77%	6.15	49	42.86 - 100
Final		89.29%	7.02	32	59.38 - 100
Grade 3	89				
Initial		93.10%	5.66	49	59.18 - 100
Final		88.48%	7.29	32	59.38 - 100
Grade 4	61				
Initial		94.08%	7.49	49	42.86 - 97.96
Final		89.50%	6.70	32	65.63 - 100
Grade 5	37				
Initial		94.87%	4.63	49	75.51 - 97.96
Final		90.88%	6.77	32	65.63 - 96.88

Table 5

Performance (% correct) by Grade on

Single-letter Consonants,

Consonant Clusters, and

Consonant Digraphs

	Grade 3 (N = 89)	Grade 4 (N = 61)	Grade 5 (N = 37)	Total sample (N = 187)
Single-letter consonants (51 items)	$\bar{X}$ = 97.00%	$\bar{X}$ = 98.47%	$\bar{X}$ = 97.30%	$\bar{X}$ = 97.54%
Range =	53.33 - 100%	53.33 - 100%	73.33 - 100%	53.33 - 100%
Consonant clusters (15 items)	$\bar{X}$ = 90.46%	$\bar{X}$ = 90.39%	$\bar{X}$ = 92.05%	$\bar{X}$ = 90.75%
Range =	66.67 - 96.08%	52.94 - 98.04%	74.51 - 96.08%	52.94 - 98.04%
Consonant digraphs (15 items)	$\bar{X}$ = 75.06%	$\bar{X}$ = 79.13%	$\bar{X}$ = 80.18%	$\bar{X}$ = 77.40%
Range =	26.67 - 86.67%	33.33 - 86.67%	46.67 - 86.67%	26.67 - 86.67%

was 95% correct or greater. One item assessing the spelling-to-sound correspondence for n had a mean percent correct score of 92.5 (because a number of students chose the foil ending with /ng/ instead of the correct response choice which ended in /n/). All other cases where mean percent correct scores were lower than 95.0 involved items assessing the three single-letter consonants (c, s, and g), which had two common sound correspondences, or items, assessing one of the two digraphs ph and ng.

For the three single-letter consonants with two common sound correspondences, four items were used to test each correspondence. Students' performances on the 24 test items assessing the consonants with variant sound correspondences are presented in Table 6. Table 6 also provides data on the percentage of students selecting each response choice. Five types of response choices were included in the test items that assessed consonants with variant correspondences: (1) a Correct response choice, (2) an Acoustically Close response choice, (3) a Visually Close response choice, (4) a response choice that was Neither Acoustically Close Nor Visually Close (Neither) to the Correct response choice, and (5) the Other Common Sound Correspondence for the Consonant. All items used a Correct, an Acoustically Close, and a Visually Close response choice. In two of the four items for each correspondence, the Neither Acoustically Close Nor Visually Close (Neither) category was used as a fourth response choice. For the remaining two items, the Other Common Sound Correspondence category was used as a fourth response choice. An inspection of Table 6

Table 6

Item Analysis of Six Variant Consonant Correspondences

for Total Subjects

(N = 187)

Target correspondence	Mean % correct	Correct response	Acoustically Close foil	Visually Close foil	Neither foil	Other Common Sound Correspondence foil
c as /k/	97	98.4	1.1	0	-	.5
		98.9	0	0	-	.5
		96.3	3.2	0	.5	-
		95.7	1.1	0	0	-
c as /s/	65	4.8	0	2.1	-	92.0
		91.4	2.7	2.7	1.6	-
		97.9	.5	0	1.1	-
		66.3	.5	23.5	-	4.8
g as /g/	95	98.9	0	1.1	0	-
		95.7	0	.5	-	2.7
		98.9	.5	.5	0	-
		87.2	.5	0	-	12.3
g as /j/	81	73.3	2.7	6.4	7.5	-
		86.6	0	3.7	1.1	-
		73.8	.5	0	-	25.1
		90.4	0	0	-	9.6
s as /s/	87	87.7	11.8	0	0	-
		94.7	.5	.5	-	3.2
		95.7	2.7	0	1.6	-
		68.4	0	2.1	-	28.3
s as /z/	73	27.3	0	2.7	-	70.1
		71.7	0	0	-	28.3
		98.4	0	0	1.1	-
		93.0	1.1	1.1	2.7	-

Note. All the totals do not add up to exactly 100% because the figures were rounded off to the nearest tenth of a percent and also because on some items a few students did not select any of the response choices.

reveals that when the Other Common Sound Correspondence category was used as a foil, students selected the Other Common Sound Correspondence response choice in percentages ranging from .5 (in both c as /k/ items) to 92.0 (in one c as /s/ item). When a variant sound correspondence for the target consonant was not present as a foil, incorrect responses were fewer and fairly evenly distributed across the three foil categories. When a variant sound correspondence was present as a foil, the mean percent correct for that item was only 73.1, compared to 92.8 on items where the variant sound was not present as a foil.

Overall performance was lowest (mean percent correct = 65.0) on items assessing the c as /s/ correspondence. In contrast, c as /k/ items yielded a mean percent correct of 97.0 for the total sample. Results for the letter-sound correspondences for s showed a mean percent correct of 73.0 for s as /z/ items and a mean percent correct of 87.0 for s as /s/ items. In analyzing performance on the two letter-sound correspondences for g in items where the variant sound was presented as a foil, the mean percent correct was 82.1 on the g as /j/ items, but was 91.4 on the g as /g/ items. In analyzing students' performance on the items for the three single-letter consonants with variant sound correspondences, performance was higher for the variant consonant correspondence in each pair that had the greater frequency of occurrence in the language (according to Venezky, Note 1).

Information on subjects' performance on the six items assessing the two digraphs ng and ph (there are three items for each digraph),

is presented in Table 7. As shown in the table, items assessing ng yielded an overall mean percent correct of 86.0 for the total group. Items assessing the ph digraph yielded an overall mean percent correct of only 72.0.

For each item assessing a digraph, the four response categories of Correct, Acoustically Close, Visually Close, and Neither Acoustically Close Nor Visually Close (Neither) were utilized. Each picture in the Visually Close category had a name which began or ended in one of the two consonant letters of the digraph. This, however, resulted in having many Visually Close foils that were also Acoustically Close, and vice versa. For example, in an item testing ph as /f/ in final position, the target synthetic word is saph. The Correct response choice is cuff and the Acoustically Close response choice is teeth. The Visually Close response choice is map, because map ends with p, one of the two consonant letters of the ph digraph. According to the Miller and Nicely (1955) confusion matrices for the perception of consonant sounds, the sound corresponding to the letter p (/p/) is sometimes confused for the sound /f/. Hence, /p/ is acoustically close to /f/, as well as visually close to ph. As shown in Table 7, when a foil category overlapped, thus creating a response that was both acoustically close and visually close, most students who responded incorrectly chose the overlapping response foil. This was the case for all six items assessing the ph as /f/ and the ng as /ŋ/ target correspondences.

Fifty-nine of the 187 children (32%) in the Consonants Study were randomly selected for testing on the two oral reading tasks. The

Table 7

Item Analysis of

Digraph Correspondences

Target correspondence	Mean % correct	Correct response	Acoustically Close foil	Visually Close foil	Overlapping Acoustic and Visual foil[a]	Neither foil
		90.4	0	7.5	-	.5
ng as /ŋ/	86	89.8	1.6	-	4.3	2.7
		79.1	-	5.3	13.9	.5
		69.5	1.1	-	23.0	3.7
ph as /f/	72	62.0	-	1.1	34.8	0
		84.5	1.6	-	10.2	1.1

Note. All the totals do not add up to exactly 100% because the figures were rounded off to the nearest tenth of a percent and also because on each item a few students did not select any of the response choices.

[a]Contained one of the two letters of the target digraph that was also acoustically close to the letter-sound correspondence for that digraph.

purpose was to ascertain whether the written test was, in fact, sensitive to the kinds of phonics errors that students are likely to make in a true reading situation. Each child was first asked to read aloud a list of the 81 synthetic words used on the written Consonants Subtest. The childrens' productive errors on target correspondences were noted by the examiners. Each student was next instructed to read aloud a 211-word poem which was written to include several words containing each of the 25 spelling-to-sound correspondences assessed in the written Consonants Subtest. Again, errors on target correspondences were noted. Results of the oral testing are presented in Table 8. Examination of this summary data indicates that students made nearly three times more errors on the written Consonants Subtest and on the oral reading of the synthetic word list than on the oral reading of the poem. Probably the benefits of context and prior knowledge of words in the poem were responsible for this difference. Children made many more errors on the spelling-to-sound correspondences when they were asked to read the list of synthetic words--words without context and about which they could have no previous knowledge. It is interesting to note, however, that agreement between errors on the Consonants Subtest and errors on the poem is not very different from the agreement between errors on the Consonants Subtest and errors on the synthetic word list. The percentage of agreement with the Consonants Subtest was 27.8 and 26.9 for the poem and word list, respectively. This provides good evidence for assuming that the written Consonants Subtest was equally sensitive to errors made on

Table 8

Summary Data on Oral Reading Tasks

and Consonants Subtest Across Grades

($\underline{N}$ = 59)

	Oral reading tasks		Consonants Subtest
	Poem	Synthetic word list	
Number of students	59	59	59
Number of students making phonics errors	11	54	59
Total number of errors	110	374	335
% of Agreement with Consonants Subtest[a]	27.8	26.9	--

[a]The percentage of agreement was calculated by determining the ratio of a student's errors common to the Synthetic Word List and Written Test over the Total Number of that Student's errors made on the Synthetic Word List alone, e.g.,

$$\% \text{ Agreement for Synthetic Word List} = \frac{\text{Number of errors common to Synthetic Word List and Written Test}}{\text{Total Number of Synthetic Word List Errors}}$$

and then summing these percentages across all students that did the oral reading tasks. The same procedure was used to calculate the percentage of agreement for the Poem and Written Test errors.

meaningful material as it was to errors made in pronouncing isolated synthetic words. For a large number of items, errors made on the oral reading of synthetic words were not made on the written Consonants Subtest. This may have resulted from the limiting nature of the foils. For example, in an item assessing the letter c as the sound /s/, a child may have read the synthetic word plice as pli/k/e, instead of as pli/s/e on the oral task. When the child came to that item on the written test, the alternate sound for c (the sound /k/) may not have been included as a foil. Because the other common sound correspondence of a consonant was used as a foil in only 50% of the items on the Consonants Subtest, the child had a better chance of responding correctly on the written item than in the productive task required by the oral reading of the synthetic word list. This foil constraint effect is suggested in Table 8 in the comparison of the total number of errors (374) made on the synthetic word list with those (335) made on the written Consonants Subtest.

Discussion

Several interesting results were obtained from examination of the data on the Consonants Subtest. Performance on consonants in initial position was superior to performance on consonants in final position, which replicates the findings in the research literature. Studies have shown that younger readers attend more to the first letters and sounds in a word than to the final letters and sounds.

In the analysis of performance in terms of the three consonant categories (single-letter consonants, consonants clusters, and consonant digraphs), a consistent pattern was revealed. All three of the grades tested scored highest on consonant clusters and lowest on consonant digraphs. While it might have been expected that performance on single-letter consonants would have been superior to performance on consonant clusters, the single-letter consonant category included items assessing variant sound correspondences for three letters (c, g, and s). It was clear from the error analysis that the inclusion of these consonants, which have more than one common sound correspondence, was responsible for the lower overall performance on items in the single-letter consonant category.

When each of the 25 spelling-to-sound correspondences was analyzed separately, the most noticeable variation in performance appeared in items where the consonants had another common sound correspondence (the consonants c, g, and s). In addition, children appeared to have experienced difficulty with two of the digraphs ph and ng. On items assessing these digraphs, children tended to pick incorrect response choices containing the first letter of each digraph, i.e., p for ph and n for ng.

Overall performance was high on the Consonants Subtest, with mean scores on all but 5 of the 25 correspondences assessed reaching at least 95% correct. Even third graders appeared to have mastered most of the Consonants Subtest items. Only a relatively small proportion of students (from all three grade levels) displayed any degree of

difficulty with consonant correspondences. Errors, when they did occur, were usually made either on consonant digraphs or on the three single-letter consonants which have other common sound correspondences (c, g, and s).

The oral reading tasks were originally planned as a measure of concurrent validity. The standard procedure for determining concurrent validity is to correlate performance on the newly devised instrument with performance on an established measure of the target skills. The survey of tests with existing phonics components revealed, however, that there is no phonics measure currently available that validly assesses childrens' phonics skills. The oral reading of the target synthetic word list appears to be a more reliable indicator of pure phonics knowledge than the oral reading of the poem. While there is no established criterion for evaluating the level of the percentage of agreement, results indicate that the written Phonics Test is sensitive to a substantial percentage of the kinds of phonics errors that children make when reading orally.

The most interesting finding of the oral reading study was that the percentage of agreement between the oral reading of the poem and performance on the written Consonants Subtest was comparable to that between the oral reading of the synthetic word list and the written Consonants Subtest. This seems to suggest that, even though children made significantly fewer errors on the poem, the errors they did make were as likely to appear on the written test as were the errors made on the synthetic word list.

DEVELOPMENT OF THE VOWELS SUBTEST

The Vowels Subtest is designed to assess 20 different spelling-to-sound correspondences. Sounds selected for testing were long vowels, short vowels, and vowel clusters.

Method

Subjects

A total of 190 third, fourth, and fifth grade pupils from the Oregon, Wisconsin Public School District participated in the present study (see Table 9). Most of these pupils had also taken the Consonants Subtest in February 1979. Third grade pupils were from Oregon Elementary School; fourth and fifth grade pupils formed a mixed-class unit in Oregon Middle School.

Stimuli

The Vowels Subtest assesses 20 different spelling-to-sound correspondences with 60 items. As in the Consonants Subtest, each item consisted of a synthetic word with the target letter(s) underlined and four response choices in picture form. Children were instructed to read the synthetic word to themselves and decide the sound of the underlined letter(s). Next they were told to circle the picture in that row whose name had a medial (or final) sound that was the same as the sound of the underlined letter(s). Figure 2 is a copy of the directions and practice items from the Vowels Subtest.

Table 9

Vowels Subtest:

Subject Population by School and Grade

(N = 190)

Schools and classrooms	Grade 3 Vowels Subtest	Grade 3 Oral reading tasks	Grade 4 Vowels Subtest	Grade 4 Oral reading tasks	Grade 5 Vowels Subtest	Grade 5 Oral reading tasks
Oregon Elementary School						
Classroom A	21	5				
Classroom B	24	5				
Classroom C	25	5				
Classroom D	22	5				
Oregon Middle School						
Classroom A			10	3	3	2
Classroom B			14	3	6	2
Classroom C			12	-	8	-
Classroom D			16	5	9	1
Classroom E			13	3	7	2
Totals	92	(20)	65	(14)	33	(7)

PHONICS: Vowels

Medial Position

In each row, look at the made-up word. Notice that there is a letter or letters underlined in the middle of that word. Read the word to yourself and decide how that letter or letters sound. Then find the picture in the row whose name contains the sound of the underlined letter or letters. Remember, you are looking for a picture whose name contains the same sound. This vowel sound may or may not be spelled with the same letters as in the made-up word. Draw a circle around the picture whose name contains the sound of the underlined letter or letters in the made-up word.

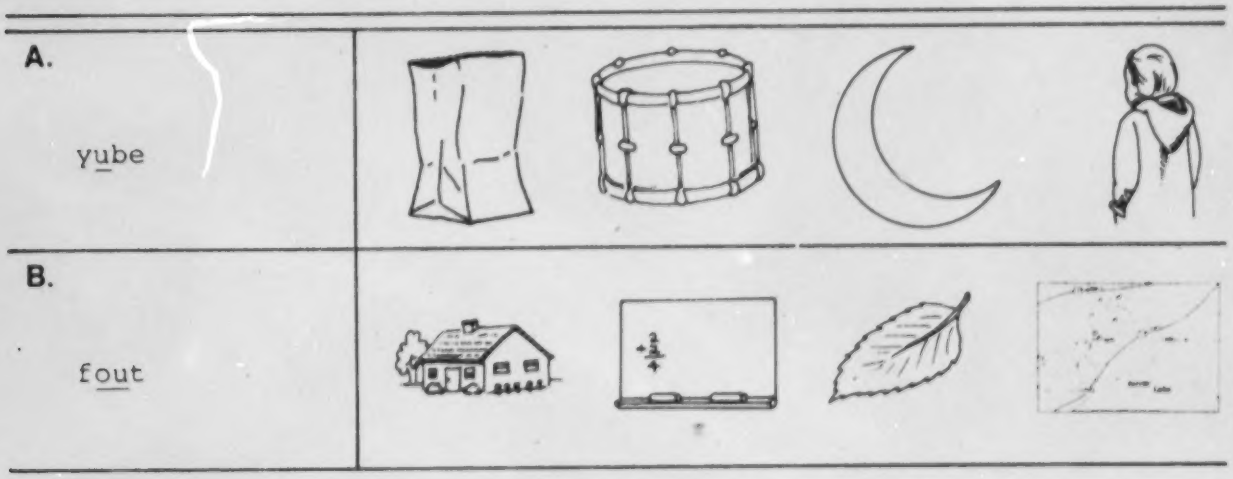


Figure 2. The directions and practice items from the first page of the Vowels Subtest.

Target vowel sounds were selected according to their frequencies of occurrence in the Venezky (Note 1) corpus of the 20,000 most common English words. Only those spelling-to-sound correspondences with total frequencies of 150 or more (with the exception of four vowel clusters, which have frequencies between 123 and 142) are included in the Vowels Subtest. These spelling-to-sound correspondences appear in one-syllable, target synthetic words that conform to phonological properties of the English language.

The response choices for each item are four pictures whose names are well known to elementary school-age children. For each target item, the four response choice categories are: (1) a Correct response choice, (2) an Acoustically Close response choice, (3) a Visually Close response choice, and (4) a Neither Acoustically Close Nor Visually Close (Neither) response choice.

Selection of Target Sounds

Long vowels. All five single-letter vowels that correspond to long vowel sounds were selected for testing. According to Venezky (Note 1), the five long vowel sounds, long a (as in cape), long e (as in mete), long i (as in hide), long o (as in bone), and long u (as in flute)[1], have spelling-to-sound correspondences with frequencies

---

[1]While the sound of u in fuse (the diphthong ju or iu) occurs more frequently than the sound of u in flute (the simple vowel u), only the latter letter-sound sound correspondence was used in the Vowels Subtest. This is because there are few one-syllable picturable words with u as in fuse that are well-known to elementary school pupils. Except for the sound correspondence for the letter u in initial position, the two sound correspondences above for u are considered to be very close.

exceeding 150. According to Kenyon and Knott's (1953), "A Pronouncing Dictionary of American English"--the source used by Venezky for the pronunciations of the 20,000 most common English words--long e and long u are always designated as simple vowels, whereas long i is always designated as a diphthong. Long a and long o, however, can be designated as a simple vowel or as a diphthong. For the purposes of this study, however, both long a and long o will be considered diphthongs. This is because the vowel sound in CVC trigrams is more commonly pronounced as a diphthong than as a simple vowel and all the target synthetic words and the picture names of the response choices used in testing vowel sounds in medial position are CVC trigrams.

Short vowels. All five single-letter vowels corresponding to short vowel sounds were selected for testing. According to Venezky (Note 1), these five short vowel sounds, short a (as in bat), short e (as in peg), short i (as in lid), short o (as in shot), and short u (as in duck), have letter-to-sound correspondences with frequencies exceeding 150. (All five short vowel sounds are designated as simple vowels.)

Vowel clusters. The 10 two-letter vowel clusters with the highest spelling-to-sound correspondence frequencies (ranging from 320 to 123, according to the Venezky tabulations) were selected for testing. Five of these clusters have simple vowel sound correspondences: ea (as in bead), ee (as in green), oo (as in spoon), au (as in taut), and ea (as in bread); the remaining five vowel clusters correspond to sounds of diphthongs: ai (as in maid), ou (as in loud),

ow (as in town), ay (as in play), and ow (as in snow).

Long vowels, short vowels, and vowel clusters are generally tested in the medial position of words which, according to the Venezky data, is the position in which they most frequently occur. The vowel cluster ay (as in play), however, occurs most frequently in final position and therefore is tested in final position only. The vowel cluster ow (as in snow) occurs frequently in medial and final position and therefore is tested in both medial and final positions.

All 55 of the items testing vowel sounds in the medial position use one-syllable picture names and target synthetic words which conform to a CVC pattern. Within any item, the final consonant sounds for the target synthetic word and the four picture names are either all voiced or all voiceless consonants. Because the final consonant sound in a CVC word affects the duration of the medial vowel (and, in the Vowels Subtest, the medial vowel sound is the sound of interest), this practice of using either all voiced or all voiceless final consonants in the response choices provides some degree of consistency for medial vowel duration.

Selection of Response Choices

Each of the 60 items in the Vowels Subtest has four response choices in picture form.

Correct response choice. This response choice is a picture whose name contains a medial or final vowel sound identical to the sound corresponding to the underlined letter(s) of the target synthetic word.

Acoustically Close response choice. This response choice is a picture whose name contains a medial or final vowel sound that is acoustically close to the sound corresponding to the underlined letter(s) in the target synthetic word. Whenever possible, data from the Peterson and Barney (1952) confusion matrix were used to determine an acoustically close sound. (The Peterson and Barney study reported perceptual confusions by adult observers under an approximately 70 decibel level for 10 simple vowel sounds, each presented 152 times by men, women, and children speakers.) To illustrate how the information from the matrix was used to determine an Acoustically Close response choice, consider the target sound short a (as in pan). The Peterson and Barney confusion matrix indicates that short a is most often confused as short e. Hence, in an item assessing the sound of short a in medial position, the target synthetic word is vad. The Acoustically Close response choice for this item is a picture of a web, because web contains the acoustically close sound of short e in medial position.

The Peterson and Barney confusion matrix does not provide confusion information for long vowel and vowel cluster sounds that are diphthongs. Hence, in these cases, speech production data (Denes & Pinson, 1973; Kenyon & Knott, 1953; Langacker, 1973) were used to construct the Acoustically Close response choice. (Evidence suggests that sounds produced in proximal positions within the oral cavity are more likely to be confused than sounds produced in more distal positions. An Acoustically Close response choice could therefore be

determined from this information.) For example, the vowel cluster or diphthong, ou, is produced from a low-central /ɑ/ to a lower-high back /ʊ/ position in the back of the oral cavity. An acoustically close vowel sound would be /ɔ/ (as in pawn), which is produced in the higher low-back position in the oral cavity--quite close to where the ou-sound is produced.[2] Therefore, in one of the items testing the sound of ou in medial position, the target synthetic word is floud. The Acoustically Close response choice for this item is a picture of saws, because saws contains the medial vowel sound /aw/.

Visually Close response choice. This response choice is a picture whose name contains the same letter as the underlined letter(s) in the target word. For short and long vowels, the sound of the letter in the picture, however, is the short or long counterpart of the sound of the underlined letter. For example, in an item testing short a the target synthetic word is vad. The Correst response choice is a picture of a pan because pan has a short a in medial position. The Visually Close response choice is a picture of a cage, because cage contains a long a in medial position. Hence, the a in cage looks like, but does not sound like, the a in vad.

For vowel clusters, we decided to use one of the two letters of the cluster for the Visually Close response choice. In an item testing ou in medial position, the target synthetic word is floud. The Visually Close response choice is a picture of a frog, because

---

[2]The terms and locations of the positions within the oral cavity for these vowel sounds are taken from Kenyon and Knott's, "A Pronouncing Dictionary of American English" (1953), p. xiii.

frog contains only one of the letters, the o of the two-letter vowel cluster ou. (Another Visually Close response choice could have been a picture of a rug, because rug contains u, the other letter in the vowel cluster ou.)

Neither Acoustically Close Nor Visually Close (Neither) response choice. This response choice is a picture whose name has neither a sound that is acoustically close, nor letters that are visually close, to the underlined letter(s) in the target synthetic word. As with the Consonants Subtest, this category was created to see whether children making errors would apply an auditory strategy (by selecting the Acoustically Close response choice), a visual strategy (by selecting the Visually Close response choice), or if they would merely guess at the correct answer. A pattern of responses having a disproportionate number in the Neither response choice category might indicate a tendency towards guessing.

For example, in the test item for short a in the target synthetic word vad, the Neither response choice is broom. The vowel sound /u/ in broom is made in the high-back position of the oral cavity. The short a in vad is produced in the low-front position of the oral cavity. Therefore, the /u/ in broom is not acoustically close to the /a/ in vad. Moreover, the oo in broom is also visually different from the single a in vad.

Procedure

For all grades, a Picture Naming Exercise was given prior to administering the Vowels Subtest of the Phonics Test. Upon completion

of the Vowels Subtest, a random sample of students also participated in two oral reading tasks.

Picture Naming Exercise. It was important to ensure that the pictures in the Vowels Subtest be identified by their designated names. Therefore, a short exercise in naming pictures preceded the administration of the Vowels Subtest. A picture was selected for the Picture Naming Exercise if it met one of two criteria: (1) the picture was created specifically for the Vowels Subtest, or (2) the picture had been included in the Picture Naming Exercise for the Consonants Subtest.

A two-page Picture Naming booklet consisting of 40 pictures was distributed to each child. The pictures were arranged in rows of four. Many of the rows were made up of pictures which shared a common feature, such as a plural name or an arrow pointing to a particular part of the picture. (Appendix C is a page from the Picture Naming booklet for the Vowels Subtest.) Students were instructed to look at one row of pictures at a time. The examiner pronounced the name of each picture in that row and the children repeated the picture name aloud. When the names of all 40 pictures had been identified through this procedure, the Picture Naming booklets were collected.

Vowels Subtest. Following the Picture Naming Exercise, each student was given a copy of the Vowels Subtest. The examiner explained to the children that they would be listening for vowel sounds in the middle and at the end of words. Next, the examiner instructed the students in the two practice items for medial vowel sounds. The

examiner then told the students to turn past the stop sign to the second to the last page in their booklets. The examiner explained that all the items up to the stop sign would be on vowel sounds in medial position; the two pages after the stop sign would be on sounds in final position. The examiner instructed the children in the one practice item for final vowel sounds. At the end of the practice period (approximately 5 minutes), students were directed to work independently on the 55 items on medial vowel sounds and on the 5 items on final vowel sounds. The children were told that if they completed all 60 items before the allocated time was up, they could review any of their work in the entire booklet.

The children took 12 to 30 minutes to complete both the medial and final vowels sections (including practice items) of the Vowels Subtest.

Oral Reading Tasks. Two oral reading tasks were individually administered to assess the concurrent validity of the Vowels Subtest. Approximately 22% of the students who took the Vowels Subtest were randomly selected to participate in the oral reading tasks. The first task required that the child read aloud a list of the 60 target synthetic words from the Vowels Subtest. As the child read the list of words, the examiner noted all pronunciation errors made on target sounds.

Following the oral reading of the word list, the child was instructed to read aloud a 347-word poem. The poem included several words containing each of the spelling-to-sound correspondences

assessed in the Vowels Subtest. The examiner noted all pronunciation errors on target correspondences. (The student's copy of the first page of the word list and the poem are presented in Appendix D.)

The total time for the administration of the oral reading tasks ranged from 3 to 10 minutes, with the word list taking approximately 1 to 5 minutes, and the poem taking approximately 2 to 5 minutes. In general, third grade pupils required more time to perform the tasks than did fourth and fifth grade pupils.

All tests were administered by staff members from the Wisconsin Research and Development Center for Individualized Schooling. Directions for the Picture Naming Exercise and the Vowels Subtest were read from an Administrator's Manual. Testing was conducted on May 22, 1979.

Results

A total of 190 students in grades three, four, and five participated in the study to evaluate the Vowels Subtest. Summary statistics on the Vowels Subtest, by grade, are presented in Table 10. Mean percent correct on the total test (60 items) was better than 75% for all three grades. An item analysis was performed to gain information on the overall reliability of the test and the relative performance of individual items within the test. The Hoyt Reliability Coefficient for the Vowels Subtest was high at .89, with a standard error of measurement of 2.85.

In addition to the item analysis, a series of t-tests were

Table 10

Vowels Subtest:

Percent Correct

Summary Statistics by Grade

Grade 3	Grade 4	Grade 5
$\underline{N}$ = 92	$\underline{N}$ = 65	$\underline{N}$ = 33
$\bar{X}$ = .7585%	$\bar{X}$ = .7808%	$\bar{X}$ = .8010%
$\underline{SD}$ = .165	$\underline{SD}$ = .118	$\underline{SD}$ = .134
$Range_{\%}$ = 18 - 100	$Range_{\%}$ = 47 - 98	$Range_{\%}$ = 52 - 100
$Range_{RS}$ = 11 - 60[a]	$Range_{RS}$ = 28 - 59[a]	$Range_{RS}$ = 31 - 60[a]

Note. Hoyt estimate of total test reliability = .89; Standard error of measurement = 2.85; Mean on total test (of 60 possible) = 46.41; and $\underline{SD}$ = 8.73.

[a]Out of a possible 60 items.

performed to determine whether there were significant differences in performance on the total test due to the sex or grade level of the students. Results of these t-tests are summarized in Table 11. There was a significant difference in scores due to sex, with female students outperforming male students (t-value = -2.10, with a 2-tailed probability of .037).

The t-tests comparing the three grades resulted in no significant differences. There were slight mean differences, with the higher grade level in each comparison attaining a somewhat higher mean percent correct score than the lower grade level of the comparison pair.

The next analysis of the Vowels Subtest data examined performance of the total sample on each of the 20 vowel spelling-to-sound correspondences assessed. The total sample data were used because no significant grade level differences in performance had been found. Table 12 presents a rank ordered listing of the 20 correspondences within each of the three general categories: short vowels, long vowels, and vowel clusters. The Vowels Subtest contained three items to assess each of the 20 vowel spelling-to-sound correspondences.

Performance on the five long vowel correspondences ranged from a mean percent correct of 85.7 for long i, to a mean percent correct of 70.3 for long u. There was a wider range of performance on the short vowel correspondences, with performance highest on short u (mean percent correct 87.7) and lowest on short o (mean percent correct 48.3).

Performance on the 10 vowel cluster correspondences ranged from

Table 11

Vowels Subtest

t-Tests for Performance Differences

Due to Sex and Grade

Variable	N	Mean	SD	SE	t-Value	df	2-Tailed probability
Male	88	.7496	.156	.017			
					-2.10*	172.15	.037
Female	102	.7941	.133	.013			
Grade 3	92	.7585	.165	.017			
					- .99	154.99	.326
Grade 4	65	.7808	.118	.015			
Grade 4	65	.7808	.118	.015			
					- .73	57.37	.466
Grade 5	33	.8010	.134	.023			
Grade 3	92	.7585	.165	.017			
					-1.46	69.08	.148
Grade 5	33	.8010	.134	.023			

*Significantly different at $p \leq .05$ level.

Table 12

Rank Ordered List of Performance Within Categories for the 20 Vowel Correspondences

Correspondence	Mean % correct
Short vowels	
u	87.7
a	86.3
i	75.0
e	70.7
o	48.3
Long vowels	
i	85.7
e	83.0
a	73.7
o	72.0
u	70.3
Vowel clusters	
ay as /ā/	94.3
ee as /ē/	93.7
ea as /ē/	91.7
oo as /ū/	87.0
ow as /ou/	84.3
ou as /ou/	80.3
ow as /ō/	73.0
ai as /ā/	68.0
au as /aw/	67.7
ea as /e/	54.7

94.3 mean percent correct on ay as long a, to 54.7 mean percent correct on ea as short e. These differences in performance on correspondences will be discussed more fully below when the analysis of error patterns and specific items are described.

A series of t-test comparisons were made to determine whether there were significant differences in performance among the three categories--short vowels, long vowels, and vowel clusters. Results are summarized in Table 13. Performance was significantly different among the three categories at the $p \leq .05$ level. As indicated in the table, performance was highest on vowel clusters and lowest on short vowels. A review of the data on the individual correspondences shown in Table 12 indicates that the low performance on the short o correspondence (mean percent correct = 48.3) lowered the overall mean for the short vowels category. The vowel clusters category, on the other hand, was inflated by three correspondences (ea as long e, ee as long e, and ay as long a) where performance topped 90%, and thus resulted in a higher overall mean for that category.

To understand why performance was so low on the short o correspondence, we reviewed item analyses for the three items assessing short o. An analysis of the foil selection for the total sample on these three items is presented in Figure 3.

For each item assessing short o, the children were presented with a target synthetic word constructed to yield the short o pronunciation. To the right of the target word were four picture response choices. One picture had a name that contained a sound acoustically

Table 13

t-Tests for Performance Differences

Due to Vowel Category

Variable	N	Mean	SD	SE	t-Value	df	2-Tailed probability
Long vowels		.7691	.195	.014			
	190				2.57*	189	.011
Short vowels		.7358	.184	.013			
Long vowels		.7691	.195	.014			
	190				-2.42*	189	.016
Vowel clusters		.7946	.145	.011			
Short vowels		.7358	.184	.013			
	190				-5.49*	189	.000
Vowel clusters		.7946	.145	.011			

*Significantly different at $p \leq .05$ level.

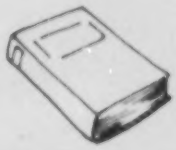
bl<u>o</u>ff				
	15.8% (book)	51.6% (knot)	28.9% (rope)	2.1% (dress)
st<u>o</u>m				
	6.3% (hood)	57.4% (bone)	35.3% (knob)	.5% (wig)
y<u>o</u>sh				
	57.9% (mop)	0% (leaf)	6.8% (hook)	33.2% (boat)

Figure 3. Item analysis of the three items assessing short <u>o</u> showing percent of students selecting responses.

close to the short o (oo as in book). Another picture had a name that contained the long sound of o (as in rope). A third picture had a name that contained a vowel sound that was neither acoustically close to the sound of short o nor visually close to the letter o (such as short e as in dress). The fourth picture had a name that contained the correct sound, a short o (as in knot). A close inspection of the errors made by students on the three short o items reveals that when students failed to correctly pronounce the vowel as a short o in the target synthetic word, they typically chose the response choice with the long o. Overall, it appears that students who made errors did not randomly guess--if this had been the case, a much higher percentage of children would have selected the response choice that was neither acoustically close nor visually close to the target spelling-to-sound correspondence (dress). Instead, students who made errors on short o items tended to choose the picture name containing the other common sound correspondence for the letter o, the long o. This was true even when the picture name had a long o sound that was spelled differently than the vowel of the target word (as in boat, where the long o sound is designated by oa). It is interesting to note that the less frequent sound /oo/ (as in hood) was not often selected, even though this sound, like the long o contains the letter o.

Performance on vowel cluster items was generally high, except for the three items assessing ea as short e. Performance on items assessing ea as long e, however, was high. This discrepancy may reflect a familiarity factor, because the frequency of ea as long e (294) is more than twice as high as the frequency of ea as short e (135).

Table 14 presents the results of an analysis of students' errors on the Vowels Subtest according to foil categories. As described in detail in the Stimulus section of this report, foil creation for the Vowels Subtest was governed by specific criteria. In addition to a Correct response choice, an item had three distractors, each containing a vowel correspondence in a picture name that was (1) Acoustically Close to the target vowel correspondence, (2) Visually Close to the target vowel, or (3) Neither Acoustically Close Nor Visually Close (Neither) to the target vowel. In some cases, the acoustic and visual distractors overlapped. When this occurred, a given test item might have one or even two response choices that contained both visual and auditory similarities to the target vowel correspondence. In considering the error analysis presented in Table 14, the reader is cautioned against drawing conclusions about this special "Both Acoustically and Visually Close" artifact.

The values in Table 14 were obtained in the following manner: Numbers of students selecting particular distractor types were summed across all test items where that distractor type was offered, and then divided by the number of items containing that distractor type, to obtain a mean percent of students selecting a particular foil or distractor. This procedure was carried out separately for the three general categories of test items: short vowels, long vowels, and vowel clusters. Inspection of these mean percent figures indicates an interesting pattern of incorrect responses. Across all three categories of vowels, students chose the Both Acoustically and Visually

Table 14

Error Analysis by Vowel Category and Foil Type

[Average Percent of Subjects (N = 189) Selecting an Incorrect Response Choice by Vowel Category]

Vowel category	Foil Type			
	Visually close	Acoustically close	Neither	Both acoustically close and visually close
Short vowels	28.3	3.9	2.0	44.6
Long vowels	35.2	7.9	1.8	17.6
Clusters	18.1	9.3	4.0	20.7
Total test	23.6	7.2	2.9	25.8

Close and the Visually Close response choices far more often than they chose the Acoustically Close or the Neither response choices.

To understand this error pattern, which was consistent across all three vowel categories, it is helpful to review the exact nature of these foils and the criteria governing their selection. For the Visually Close response choice, we decided to use pictures whose names contained the same vowel letter as in the target synthetic word (or, for clusters, one of the two vowel letters in the target cluster). In the distractor name, however, the letter would correspond to a different sound. For example, if short a were the target sound, the Visually Close response choice might be a picture of a cave. A child using a visual or spelling strategy might select the picture of a cave, because cave contains the same letter a that is in the target synthetic word. On the other hand, if the student was not familiar with the rules governing pronunciation of the short a in the target synthetic word and therefore pronounced the vowel sound as long a, cave would be a logical response. Both of these approaches could have led students to select the Visually Close response choice cave and contributed toward making the Visually Close and the Both Acoustically and Visually Close response choices the most popular distractors when errors were made.

Because it is not possible to determine which features of the Visually Close and the Both Acoustically and Visually Close response choices led students to choose these responses, conclusions cannot be drawn as to the relative strengths of the visual versus the acoustic features within these foils. It is reasonable to say, however, that

children who failed to correctly decide target correspondences appeared to have been strategic, rather than random, in their selection of response choices.

Three additional analyses of the raw score data of the Vowels Subtest were performed. First, data were analyzed to determine whether items ending in a voiced or a voiceless consonant affected students' performance on items assessing vowels in medial position. Because voiced consonant endings generally lengthen the duration of the preceding vowel sound, such endings might enhance students' perception of vowel sounds and thus skew performance. This was not the case in the Vowels Subtest, however. Mean performance scores for items containing all voiced consonant endings were nearly identical to mean scores for items containing all voiceless consonant endings.

A second analysis was performed to determine whether performance would differ on items where the spelling of the vowel sound in the name of the correct picture differed from the vowel spelling in the target synthetic word, versus those items where the spelling was identical for the vowel sound in the picture and in the target synthetic word. Mean percent correct on items representing these two cases differed. Items with identical spellings had a mean percent correct of 81.1, whereas items with different vowel spellings had a mean percent correct of 71.3. Because vowel correspondences with relatively smaller frequencies were somewhat over-represented by items that were spelled differently, however, this mean percent correct discrepancy was not considered important.

The third analysis of the Vowels Subtest data involved scores collected from a subset of the total sample. Forty-two of the 190 children (22%) in the total sample were randomly selected for individual productive testing on two oral reading tasks. The purpose was to obtain information on the concurrent validity of the written Vowels Subtest. The question of interest was whether the group administered written test was, in fact, sensitive to the kinds of phonic errors on vowels that students are likely to make in a typical reading situation.

Children participating in the oral tasks were first asked to read aloud a list of the 60 synthetic words used on the written Vowels Subtest and their productive errors on target correspondences were noted by examiners. The children were next instructed to read aloud a 347-word poem which included several words for each of the 20 spelling-to-sound correspondences assessed in the written test. Again, errors on target correspondences were recorded.

Results of the error analysis on the oral testing data are presented in Table 15. Examination of the summary data in Table 15 indicates that students made nearly three times as many errors on the written test as on the poem and nearly three times as many errors on the synthetic word list as on the poem. It should be noted that the poem contained only words which would be well known to students in the grade levels tested. The combination of recognition of the poem's words as sight words and the benefits of reading words in context probably accounts for the higher performance on the oral reading of

Table 15

Summary Data on Oral Reading Tasks and

Vowels Subtest Across Grades

($\underline{N}$ = 42)

	Oral reading tasks		Vowels Subtest
	Poem	Synthetic word list	
Number of students	42	42	42
Number of students making phonics errors	27	39	41
Total number of errors made	137	398	465
% of Agreement with Vowels Subtest[a]	41.6	52.8	--

[a]The percentage of agreement was calculated by determining the ratio of a student's errors common to the Synthetic Word List and Written Test over the Total Number of that student's errors made on the Synthetic Word List alone. For example,

$$\% \text{ Agreement for Synthetic Word List} = \frac{\text{Number of errors common to Synthetic Word List and Written Test}}{\text{Total Number of Synthetic Word List Errors}}$$

Individual percentage of agreement scores were then summed across subjects. The same procedure was used to calculate the percentage of agreement for the Poem and the Written Test errors.

the poem. The percentage of agreement between the poem and the written test, and between the synthetic word list and the written test, however, are not discrepant. Even though students were less likely to make errors on target vowel correspondences while reading the poem, the errors they did make were also made on the written test an average of 41.6% of the time. The percentage of agreement for errors made on the synthetic word list and on the written test was only slightly higher, with errors on vowel correspondences agreeing on the average of 52.8% of the time. The percentage of agreement was calculated by using the ratio presented in Table 15. Although there may be alternate ways to calculate agreement, the ratio used was based on the consideration that the risk would be greater from an educational viewpoint if the written test were not sensitive to possible confusions on phonics generalizations (i.e., if it pointed out relatively few areas of possible problems). The risk would be less if the test were relatively oversensitive to possible confusion areas (i.e., if it identified more areas of possible problems) which the classroom teacher could later rule out as not troublesome through informal checks.

Discussion

The most interesting results of the Vowels Subtest relate to the analysis of error patterns made when students had not mastered a correspondence. Even when making errors, children appeared to have been strategic in selecting their response choices.

Several issues that were of concern during item construction were analyzed. Items ending in voiced versus voiceless consonants were

evaluated to determine whether such consonant manipulation would affect vowel perception, making certain items (namely, those ending in voiced consonants) easier to perceive than others. In fact, no such effect was obtained. We decided, however, to maintain the consistency within items of having the target synthetic word and the four picture names end in either all voiceless or all voiced consonants.

The spelling of the target vowel correspondences in the synthetic words versus the spelling for the vowel sound in the correct response picture name was also examined. Performance was the same for items with identical spellings and items with discrepant spellings.

Although much care was taken to ensure the distinctiveness of each response foil, overlapping features were sometimes unavoidable. This is due to the large number of sounds generated by relatively few vowel letters and the similarities of many of the vowel sounds. While it is clear that the children chose these overlapping foils in fairly high percentages, the strategies they used in their selections cannot be explained with certainty. It can be speculated that the relatively high numbers of students selecting the Visually Close and the Both Acoustically and Visually Close response choices represent a group that when they mispronounce a vowel, do so by saying the vowel's most frequent sound counterpart. In addition, some pupils may mentally "spell" the picture names of the response foils, and then "match" the vowel spelling of the picture name to that found in the target synthetic word. While neither explanation is clearly dictated by data, it is certain that students who failed to master the 20 vowel spelling-to-

sound correspondences used some kind of strategic approach in responding to items.

Results of the error analysis involving the oral testing indicated that students made more than three times as many errors on the written test as on the poem and nearly three times as many errors on the synthetic word list as on the poem. The easy vocabulary of the poem, as well as the advantage of being able to read the words in context, probably account for this discrepancy. Of more interest was the relatively even percentage of agreement between the kinds of phonics errors made on the poem and the written test and on the synthetic word list and the written test (41.6 and 52.8%, respectively).

The percentage of agreement was calculated by dividing the errors in common on the oral and written tasks by the total number of errors made on the oral task. Although there are alternate ways to calculate agreement, this ratio was used because it would be the most sensitive to possible confusions in phonic generalizations. The Phonics Test will be of greater diagnostic value if it identifies all possible phonics problems (which, if necessary, could be confirmed through an informal check) than if the test is not sensitive to possible errors and is therefore not thorough in pointing out all of a student's phonic weaknesses.

SUMMARY AND FUTURE DIRECTIONS

Results of item analyses from the administration of the spring 1979 Phonics Test suggested some interesting speculations about the

nature of phonics processing. Overall, performance on both the Consonants and Vowels Subtests was high. The results of the Consonants study indicated that by third grade children have mastered most of the frequently occurring letter-sound correspondences in our language. Exceptions included items assessing two of the digraphs (ph and ng). Children who made errors on these tended to select distractors containing either p (for the ph digraph) or n (for the ng digraph)--that is, the first of the two letters comprising each digraph. Performance also dropped on the three single-letter consonants which have other common sound correspondences (c, s, and g). Predictably, when children made errors on items assessing these letters with variant correspondences, they selected responses containing the other common sound for each letter (i.e., g as /j/ instead of as /g/, c as /s/ instead of as /k/, and s as /z/ instead of as /s/).

In general, performance was highest on consonant clusters (mean percentage = 97.54), second highest on single-letter consonants (mean percentage = 90.75), and lowest on consonant digraphs (mean percentage = 77.4). Between-grade differences were not significant, with third, fourth, and fifth grade pupils averaging 91.3, 92.3, and 93.3% correct, respectively. For all grades, performance was higher on consonants assessed in initial position than on those assessed in final position. Agreement on the kinds of phonics errors made orally and on the written test was 27.8% for the poem and 26.9% for the synthetic word list.

In the Vowels Subtest mean performance, while not as high as in the Consonants Subtest, still exceeded 75% for all three grades.

There were no significant differences between grades, although significant sex differences were obtained--with girls outperforming boys across the three grade levels.

A series of t-test comparisons revealed that there were significant differences in performance among the three categories of vowel items--short vowels, long vowels, and vowel clusters. Performance was highest on vowel clusters (mean score 79.5%) and lowest on short vowels (mean score 73.6%).

An analysis of the kinds of errors made on the Vowels Subtest revealed that, in general, children who made errors were strategic and selected response choices that were either acoustically or visually close to the target letter(s) or sound, or which represented another common sound correspondence for the target letter.

Finally, analyses designed to reveal possible sources of error relating to the construction of the Vowels Subtest were performed. These analyses involved comparing (1) items ending in voiced consonants versus items ending in voiceless consonants, (2) items where the spelling of a target sound matched the spelling of the name of the correct response picture with items in which the spelling of the target sound in the synthetic word differed from the spelling of the sound in the correct picture name, and (3) a subsample of children's performance on two oral reading tasks with performance on the written Vowels Subtest.

The above analyses yielded three major conclusions. First, the children did not perform differentially on vowel items that ended in

all voiced consonants versus items that ended in all voiceless consonants. Second, spelling differences of vowel sounds did affect performance. Many relatively unfamiliar correspondences were overrepresented among the "spelled differently" items, however, thus clouding interpretation of this finding. Finally, far fewer phonics errors were made on the oral task with the poem than on either the oral task with the synthetic word list or the written test. Nevertheless, the percentage of agreement between each oral task and the written test was extremely close--41.6% for the poem and the written test, and 52.8% for the synthetic word list and the written test.

As a result of these analyses, several changes are being considered for the winter 1979-80 Phonics Test. One issue of primary interest is the relative contribution of children's auditory and visual strategies in selecting response choices. Because the Acoustically Close and the Visually Close response choices overlapped in a large number of items, a clear-cut analysis of error patterns across the entire test was difficult.

In an attempt to eliminate some of the overlapping between response categories, the rules for creating response choices will be revised. As in the spring 1979 version of the test, all items will have four response choices, including a Correct response choice and a Neither response choice. The remaining two response choices will be from the following three categories: Acoustically Close (based on the Miller and Nicely confusion matrices for consonants, and the Peterson and Barney confusion matrix for vowels), Visually Close (based on the Bouma confusion matrix for lower-case letters), and the Other Common

Sound of the Target Correspondence (for example, long a if the target correspondence is short a, and c as /s/ if the target correspondence is c as /k/). For some items, however, an overlapping between response choice categories will be unavoidable, as, for example, in the case of m. There is no Other Common Sound Correspondence for m and, therefore, the response choices will be from the two categories Acoustically Close and Visually Close. The sound that is closest acoustically to the target sound /m/ is /n/. Similarly, the letter n is the letter that is visually closest to the target letter m. Hence, the target correspondence in the picture names in both response choices will inevitably overlap.

To more clearly delineate the relative contributions of students' auditory and visual strategies, consideration is being given to examining only those items in which none of the response choices overlap when analyzing the data for error patterns. Response foils for this analysis will be from the categories of Acoustically Close, Visually Close, Other Common Sound of the Target Correspondence, and Neither.

A second issue involves assessing younger children's performance on the Phonics Test. Because overall performance on both the Vowels and Consonants Subtests was extremely high, and because we are interested in gaining a better understanding of the developmental nature of phonics skills acquisition, students in the primary grades will participate in the next administration of the tests. The spring 1979 Phonics Test was administered to pupils in third, fourth, and fifth

grades; the revised Phonics Test will be administered to second and third grade children in winter 1979-80.

Another issue raised during test administration concerned the effect that misnamed pictures might have had on children's performance. While the children participated in the Picture Naming Exercise immediately prior to the actual administration of the test, there was still uncertainty as to whether children recalled or used the intended labels for the pictures. Moreover, the Picture Naming Exercise, with its separate set of directions, required an additional phase of test administration. Therefore, in the next administration of the test the examiner will read aloud the names of the picture response choices to students. This alteration in procedure will eliminate the need for the Picture Naming Exercise and will control for consistency in labeling pictures. An additional benefit is that students will be paced throughout the test. Because students must still read (decide) the target synthetic word themselves, the decoding nature of the task will be preserved.

The most significant issue concerning changes in the next version of the Phonics Test involves the assessment of more letter-sound correspondences. To make the test more global, without adding substantial time to the test, the number of items used to assess each correspondence will be decreased from three or four to two. For example, on the spring 1979 Phonics Test, 15 single-letter consonant correspondences were randomly selected for assessment from a category total of 23. On the new version of the Phonics Test, in which there

will be only two items per correspondence, all 23 single-letter consonants can be assessed with only 46 items--five items fewer than on the earlier version. Table 16 presents a breakdown of the number of items assessed in the spring 1979 Phonics Test for each consonant and vowel category, and of the projected number of items being considered for inclusion in the winter 1979-80 Phonics Test. As Table 16 indicates, the new version will assess 73% more correspondences in only 11% more items.

In winter 1979-80, the revised version of the Phonics Test will be pilot-tested on several classes of primary school children. Appropriate revisions will be made following analysis of the data.

The Phonics Test is one of three major components of a Word Identification Test battery. The total battery will be administered in April 1980 to several hundred elementary school children. The reading subtest of the Metropolitan Achievement Tests (Durost et al., 1970), a standardized measure of comprehension, will also be administered in order to obtain correlationary information. All-ways correlations will be run between each of the components (Phonics, Structure, and Context) of the total Word Identification Test battery and the comprehension sections of the Metropolitan Achievement Tests.

We believe that the information gained through the administration of the test battery will provide insights about the relationships of the various components of word identification to comprehension. Ultimately, this knowledge will have practical application for instruction in the classroom.

Table 16

Comparison Between the Spring 1979 Phonics Test and the Proposed Revised Phonics Test for Number of Correspondences Assessed and Number of Items per Category

	Spring 1979 Phonics Test		Proposed Revised Phonics Test	
	No. of correspondences assessed	No. of items (3 or 4 per correspondence)	No. of correspondences assessed	No. of items (2 per correspondence)
Consonants Subtest				
Single-letter consonants	15	51	23	46
Consonant clusters	5	15	21	42
Consonant digraphs	5	15	5	10
Total	25	81	49	98
Vowels Subtest				
Long vowels	5	15	5	10
Short vowels	5	15	5	10
Other single-letter vowels	-	-	4	8
Vowel clusters	10	30	15	30
Total	20	60	29	58
Total Phonics Test	45	141	78	156

Reference Notes

1. Venezky, R. L. A computer program for deriving spelling to sound correlations. Master's Thesis, Cornell University, 1962. Published in part in H. Levin et al., A basic research program on reading. Ithaca, NY: Cornell University, 1963.

2. Shriberg, L. D. Personal communications, June 1978 - December 1978.

References

Bond, G. L., Balow, B., & Hoyt, C. J. The silent reading diagnostic tests. Chicago: Lyons & Carnahan, Division of the Rand McNally College Publishing Co., 1970.

Botel, M. Botel reading inventory. Chicago: Follett Publishing Co., 1961.

Bouma, H. Visual recognition of isolated lower-case letters. Vision Research, 1971, 11, 459-474.

Brown, G. M., & Cottrell, A. B. California phonics survey. Monterey, CA: McGraw-Hill, 1963.

California achievement tests. Monterey, CA: McGraw-Hill, 1977.

Denes, P. B., & Pinson, E. N. The speech chain: The physics and biology of spoken language. New York: Anchor Press, 1973.

Durkin, D., & Meshover, L. Phonics knowledge survey. New York: Teachers College Press, Columbia University, 1964.

Durost, W. N., Bixler, H. H., Wrightstone, J. W., Prescott, G. A., & Balow, I. H. Metropolitan achievement tests, Form F. New York: Harcourt Brace Jovanovich, Inc., 1970.

Johnson, D. D., Pittelman, S. D., Schwenker, J., Shriberg, L. K., & Morgan-Janty, C. Relationships between word identification skills and reading comprehension of elementary school children (Technical Report No. 490). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1978.

Kenyon, J. S., & Knott, T. A. A pronouncing dictionary of American English. Springfield, MA: G. & C. Merriam Co., 1953.

Kress, R. A. In O. K. Buros (Ed.), The seventh mental measurements yearbook (Vol. II). Highland Park, NJ: Gryphon Press, 1972.

Langacker, R. W. Language and its structure. New York: Harcourt Brace Jovanovich, Inc., 1973.

Madden, R., Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. Stanford achievement test. New York: Harcourt Brace Jovanovich, Inc., 1970-1974.

Massaro, D. W. Understanding language. New York: Academic Press, 1975.

Miller, G. A., & Nicely, P. E. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 1955, 27, 338-352.

Otto, W., Miles, P. J., Kamm, K., & Stewart, D. M. The Wisconsin Design for reading skill development. Minneapolis: National Computer Systems, Inc., 1972.

Peterson, G. E., & Barney, H. L. Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 1952, 24, 175-184.

Prescriptive reading inventory. Monterey, CA: CTB/McGraw-Hill, 1972 and 1976.

Skills monitoring system for reading and word identification. New York: Harcourt Brace Jovanovich, Inc., 1975.

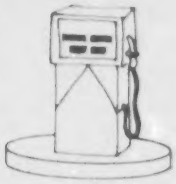
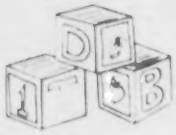
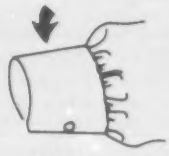
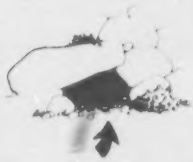
Venezky, R. L. The structure of English orthography. The Hague: Mouton, 1970.

Appendix A

A page from the Picture Naming booklet

for the Consonants Subtest

Picture Naming

1.	2. 	3.	4.
5. 	6.	7.	8.
9.	10.	11.	12. 
13.	14.	15.	16.
17.	18. 	19.	20.

Appendix B

The subject's copy of the first page of the word list from the oral reading task for the Consonants Subtest

The student's copy of the poem from the oral reading task for the Consonants Subtest

Below is a list of "made-up" words. Please read each word aloud.

1.	cade	15.	thelp
2.	vode	16.	wift
3.	dem	17.	joast
4.	troon	18.	wame
5.	welp	19.	tasp
6.	speat	20.	chone
7.	chort	21.	brate
8.	gade	22.	cluff
9.	spafe	23.	cuzz
10.	jild	24.	cete
11.	clote	25.	traff
12.	kisp	26.	broop
13.	brask	27.	goap
14.	cobe	28.	baff

"If You Weren't You"

Have you ever thought of things
From a different point of view?
Of how life would be
If you weren't you?

Would you like to play a game with me?
Just for fun?
Let's pretend we're something different.
Here's how the game is done.

If you were a star shining in space
You'd never have to comb your hair
Or wash your hands and face.

If you were a king, very long ago
You could never ever chew gum
Or watch a TV show.

If you were a tree in a field of flowers,
You'd have no use for clocks or kites
Or bathtub toys or towers.

If you were a bride with a train of silk,
You might pose for a photograph
While drinking chocolate milk.

If you were a spider
Living in a rug,
You might build a g[illegible]t web
To trap the biggest bug.

If you were a clown with a bright red nose,
You might jump through a drum
Or spread jam with your toes.

If you were a book on birds and fishes,
One page might tell why both of these
Think worms make tasty dishes.

But I am I, and you are you,
And so my friend
Our game is through.

Appendix C

A page from the Picture Naming booklet

for the Vowels Subtest

Picture Naming

86

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20

Appendix D

The subject's copy of the first page of the word list from the oral reading task for the Vowels Subtest

The student's copy of the poem from the oral reading task for the Vowels Subtest

Below is a list of "made-up" words. Please read each word aloud.

1. bap
2. shuff
3. prown
4. stuke
5. roove
6. drave
7. floud
8. naud
9. neek
10. prube
11. teap
12. vad
13. duss
14. fide
15. frain
16. treen
17. fream
18. yede
19. blute
20. trowse
21. veem
22. maith
23. spash
24. vede
25. bloff
26. zope
27. taun
28. zoon

Gabe and I are the best of friends.
Through boom and bust we'll be friends to the end.
But there was one day when we almost parted,
I'll tell you the story. Here's how it all started. . .

We were off on a journey to find the mouth
Of the Blue Moon River (it's way down south).
On a nice June day we were packing our jeep,
We'd had breakfast early and had finished our sleep,

When suddenly I saw what I thought was a crow
Stealing Gabe's old red hat from our campsite below.
That brainy old bird took the hat in its beak
And flew over some trees with it---that old sneak!

Since Gabe had just gone to the lake for some water,
I ran through the trees on the trail of the robber,
I'd find where he'd gone and be back in a minute,
And give Gabe his hat before he had missed it.

(At least those were the thoughts that I thought at the scene.
How mistaken I was gives my story its theme.)

The old crow had flown with the hat to his nest,
And I thought I could see him, shredding its crest.
I'd sat down on a rock to come up with a plan
When I heard Gabe shouting, so back I ran.

Gabe was red in the face and mad as a hen.
He demanded to know where the heck I had been.
Seems he'd finished the chores, all the cleanup and packing,
And he yelled that when work called I always was lacking.

Well I tried to explain 'bout his hat and the crow,
But his face, it got taut and his anger did grow. . .
Since his red hat was sitting on top of his head,
He just naturally thought I was lying instead.

Now I still don't know what it was that crow took,
But it taught me to give things a second hard look.
If I think I've seen one thing, I'll look twice again,
'Cause a red hat that isn't can lose you a friend.